

The 2020 Census Disclosure Avoidance System

Michael Hawes

Senior Advisor for Data Access and Privacy
Research and Methodology Directorate
U.S. Census Bureau

May 26, 2021

Shape
your future
START HERE >

United States[®]
Census
2020

Acknowledgements

This presentation includes work by the Census Bureau's 2020 Disclosure Avoidance System development team, Census Bureau colleagues, and our collaborators, including: John Abowd, Tammy Adams, Robert Ashmead, Craig Corl, Ryan Cummings, Jason Devine, John Fattaleh, Simson Garfinkel, Nathan Goldschlag, Michael Hawes, Michael Hay, Cynthia Hollingsworth, Michael Ikeda, Kyle Irimata, Dan Kifer, Philip Leclerc, Ashwin Machanavajjhala, Christian Martindale, Gerome Miklau, Claudia Molinar, Brett Moran, Ned Porter, Sarah Powazek, Vikram Rao, Chris Rivers, Anne Ross, Ian Schmutte, William Sexton, Rob Sienkiewicz, Matthew Spence, Tori Velkoff, Lars Vilhuber, Bei Wang, Tommy Wright, Bill Yates, and Pavel Zhurlev.

For more information and technical details relating to the issues discussed in these slides, please contact the author at michael.b.hawes@census.gov.

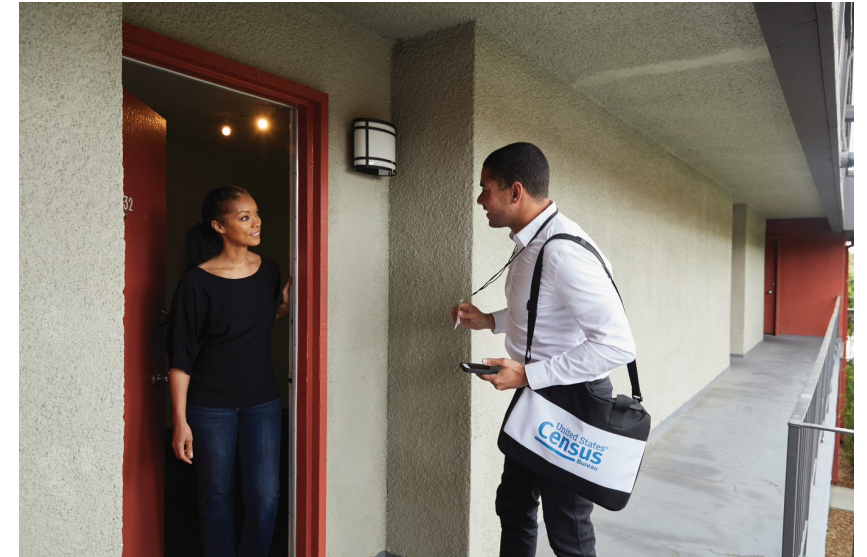
Any opinions and viewpoints expressed in this presentation are the author's own, and do not represent the opinions or viewpoints of the U.S. Census Bureau.

The statistics included in this presentation have been cleared for public dissemination by the Census Bureau's Disclosure Review Board (CBDRB-FY20-DSEP-001, CBDRB-FY20-281, and CBDRB-FY20-101).

Our Commitment to Privacy and Confidentiality

Data stewardship is central to the Census Bureau's mission to produce high-quality statistics about the people and economy of the United States.

Our commitment to protect the privacy of our respondents and the confidentiality of their data is both a legal obligation and a core component of our institutional culture.



Upholding our Promise: Today and Tomorrow

We cannot merely consider privacy threats that exist today.

We must ensure that our disclosure avoidance methods are also sufficient to protect against the threats of tomorrow!



The Census Bureau's Privacy Protections Over Time

Throughout its history, the Census Bureau has been at the forefront of the design and implementation of statistical methods to safeguard respondent data.

Over the decades, as we have increased the number and detail of the data products we release, so too have we improved the statistical techniques we use to protect those data.



The Privacy Challenge

Every time you release any statistic calculated from a confidential data source you “leak” a small amount of private information.

If you release too many statistics, too accurately, you will eventually reveal the entire underlying confidential data source.

Dinur, Irit and Kobbi Nissim (2003) “Revealing Information while Preserving Privacy” PODS, June 9-12, 2003, San Diego, CA



The Growing Privacy Threat

More Data and Faster Computers!

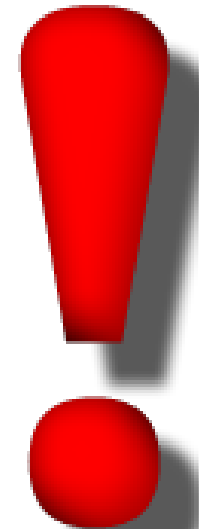
In today's digital age, there has been a proliferation of databases that could potentially be used to attempt to undermine the privacy protections of our statistical data products.

Similarly, today's computers are able to perform complex, large-scale calculations with increasing ease.

These parallel trends represent new threats to our ability to safeguard respondents' data.

The Census Bureau's Decision

- Advances in computing power and the availability of external data sources make database reconstruction and re-identification increasingly likely.
- The Census Bureau recognized that its traditional disclosure avoidance methods are increasingly insufficient to counter these risks.
- To meet its continuing obligations to safeguard respondent information, the Census Bureau has committed to modernizing its approach to privacy protections.



Disclosure Avoidance

Disclosure avoidance methods seek to make reconstruction and re-identification more difficult, by:

- Reducing precision
- Removing vulnerable records, or
- Adding uncertainty

Commonly used (legacy) methods include:

- Complementary suppression
- Rounding
- Top/Bottom coding of extreme values
- Sampling
- Record swapping
- Noise injection

Problem #1 – Impact on Data

All statistical techniques to protect privacy impose a tradeoff between the **degree of privacy protection** and the resulting **accuracy of the data**.

Swap rates, noise injection parameters, cell suppression thresholds, etc. determine this tradeoff.

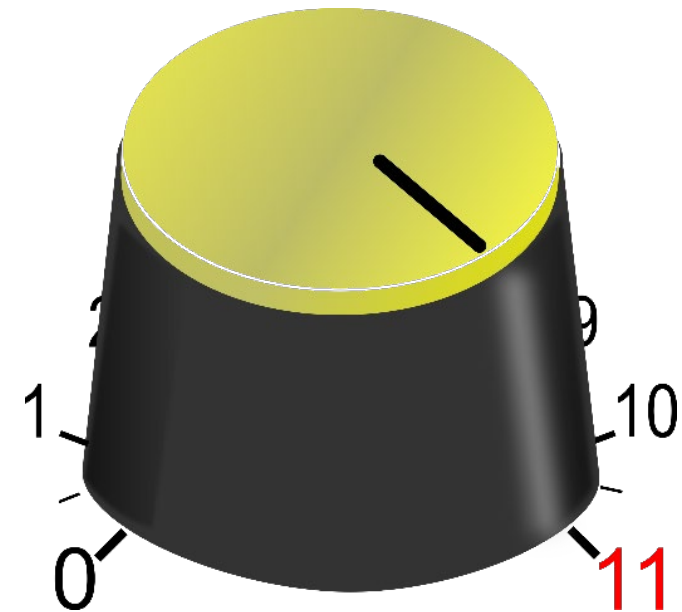


Problem #2 – How much is enough?

Legacy disclosure avoidance methods provide little ability to quantify privacy protections.

When faced with rising disclosure risk, disclosure avoidance practitioners adjust their implementation parameters.

BUT, this is largely a scattershot solution that over-protects some data, while often under-protecting the most vulnerable records.



Differential Privacy

DP is not a disclosure avoidance “method” as much as it is a framework for defining and then quantifying privacy protection.

Every individual that is reflected in a particular statistic contributes towards that statistic’s value.

Every statistic that you publish “leaks” a small amount of private information.

DP as a framework allows you to assess each individual’s contribution to the statistic, and to measure (and thus, limit) how much information about them will leak.



Differential Privacy

When combined with noise injection, DP allows you to precisely control the amount of private information leakage in your published statistics.

- Infinitely tunable – parameter “dials” can be set anywhere from perfect privacy to perfect accuracy.
- Privacy guarantee is mathematically provable and future-proof.
- The precise calibration of statistical noise enables optimal data accuracy for any given level of privacy protection.*

*Absent post-processing requirements, which can introduce error independent of that needed to protect privacy.



Privacy vs. Accuracy

The only way to absolutely eliminate all risk of re-identification would be to never release any usable data.

Differential privacy allows you to quantify a precise level of “acceptable risk,” and to calibrate your disclosure avoidance mechanism to a precise point on the privacy/accuracy spectrum for the resulting data.

Providing accurate data



Safeguarding individual privacy

Data Quality | Bnae Kegouqe
Dada Qualitg | Vrkk Jzcfkdy
Data Qaality | Dncb PrhvBl
Dzte Qvality | Dncb Prtnavy
Dfha Quapyti | Tgta Ppijacy
Tgta Qucjity | Dfha Pnjvico
Dncb Qhulitn | Dzhe Njivaci
Ntue Quevdto | Dzte Privacy
Vrkk Zuhnvry | Dada Privacg
Bnaq Denorbe | Data Privacy

Establishing a Privacy-loss Budget

This measure is called the “Privacy-loss Budget” (PLB) or “Epsilon.”

$\epsilon=0$ (perfect privacy) would result in completely useless data

$\epsilon=\infty$ (perfect accuracy) would result in releasing the data in fully identifiable form



Epsilon

Implications for the 2020 Census

The modernization of our privacy protections using a differential privacy framework does not change the constitutional mandate to apportion the House of Representatives according to the actual enumeration.

As in 2000 and 2010, the Census Bureau will apply privacy protections to the P.L. 94-171 redistricting data, and all subsequent 2020 Census data products.

Demonstration Data

- Since October 2019, the Census Bureau has been periodically releasing demonstration data products (using 2010 Census data) for data user evaluation.
- The first four of these sets of demonstration data (October 2019, May 2020, September 2020, November 2020) used a conservative global PLB set by DSEP for the October 2019 Demonstration Product, in order to evaluate algorithmic improvements.
- ***The 2020 Census Data Products will not be held to this fixed PLB.***
- On April 28, 2021 we released another set of Privacy-Protected Microdata Files (PPMFs) and Detailed Summary Metrics using a different global PLB ($\epsilon=12.2$) that more closely approximates the level of PLB that the DSEP will be considering for the 2020 Census redistricting data files.
- In September, we plan to release a final set of PPMFs using the actual production code and settings that will be used for the 2020 Census redistricting data files.

How to Submit Feedback

The changes in the [April 2021 PPMFs](#) data set reflect the cumulative feedback received from the data user community throughout the development process. We look forward to feedback from data users on this [new demonstration product](#). Your input will inform the Census Bureau's June 2021 final decision on the PLB and on the 2020 Census redistricting data parameters. **The deadline to submit feedback is May 28, 2021.**

**** Please send comments to 2020DAS@census.gov with the subject line "April 2021 Demonstration Data."**

Particularly useful feedback would describe:

- **Fitness-for-use:** Based on your analysis, would the data needed for your applications (redistricting, Voting Rights Act analysis, estimates, projections, funding data sets, etc.) be satisfactory?
 - How did you come to that conclusion?
 - If your analysis found the data to be unsatisfactory, how incrementally would accuracy need to change to improve the use of the data for your required or programmatic use case(s)?
 - Have you identified any improbable results in the data that would be helpful for us to understand?"
- **Privacy:** Do the proposed products present any confidentiality concerns that we should address in the DAS?
- **Improvements:** Are there improvements you've identified that you want to make sure we retain in the final design? Be specific about the geography and error metric for the proposed improvement.

Shape
your future
START HERE >

United States[®]
Census
2020

Stay Informed:
Subscribe to the 2020 Census Data
Products Newsletters

*Search “Disclosure Avoidance” at www.census.gov

2020 Census Population Counts for Apportionment are Now Available

// Census.gov > [2020 Census Research, Operational Plans, and Oversight](#) > [Process](#) > [Disclosure Avoidance Modernization](#) > [2020 Census Data Products Newsletters](#)



2020 Census Data Products Newsletters

Sign up for news and information about 2020 Census Data Products and the implementation of the new Disclosure Avoidance System.

[SIGN-UP FOR NEWSLETTERS](#)

Past Issues:

April 28, 2021

New DAS Update Meets or Exceeds Redistricting Accuracy Targets

April 19, 2021

New Demonstration Data Will Feature Higher Privacy-loss Budget

April 07, 2021

Meeting Redistricting Data Requirements: Accuracy Targets

February 23, 2021

The Road Ahead: Upcoming Disclosure Avoidance System Milestones

February 03, 2021

New DAS Phase: Optimizing Tunable Elements

November 25, 2020

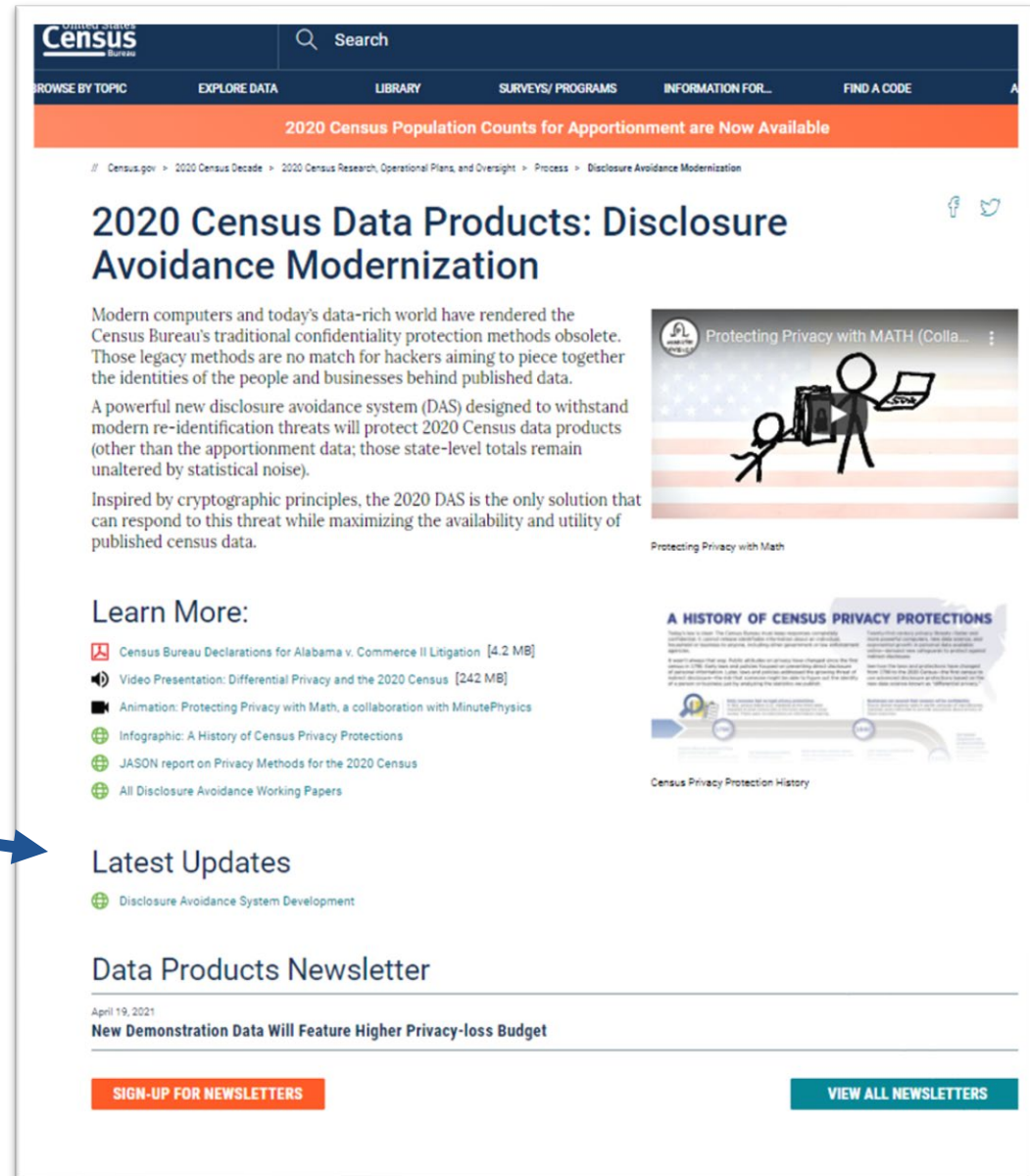
Invariants Set for 2020 Census Data Products

Stay Informed: Visit Our Website

*Search “Disclosure Avoidance” at www.census.gov

Latest Updates

 [Disclosure Avoidance System Development](#)



The screenshot shows the top navigation bar of the Census Bureau website with a search bar and menu items: BROWSE BY TOPIC, EXPLORE DATA, LIBRARY, SURVEYS/ PROGRAMS, INFORMATION FOR..., and FIND A CODE. Below the navigation is an orange banner with the text "2020 Census Population Counts for Apportionment are Now Available". The main content area features the article title "2020 Census Data Products: Disclosure Avoidance Modernization" with a breadcrumb trail: // Census.gov > 2020 Census Decade > 2020 Census Research, Operational Plans, and Oversight > Process > Disclosure Avoidance Modernization. The article text discusses the obsolescence of traditional confidentiality methods and the introduction of a new Disclosure Avoidance System (DAS) designed to withstand modern re-identification threats. It includes a sub-header "Learn More:" with a list of resources: "Census Bureau Declarations for Alabama v. Commerce II Litigation [4.2 MB]", "Video Presentation: Differential Privacy and the 2020 Census [242 MB]", "Animation: Protecting Privacy with Math, a collaboration with MinutePhysics", "Infographic: A History of Census Privacy Protections", "JASON report on Privacy Methods for the 2020 Census", and "All Disclosure Avoidance Working Papers". To the right of the text is an infographic titled "Protecting Privacy with MATH (Colla...)" and another titled "A HISTORY OF CENSUS PRIVACY PROTECTIONS". At the bottom of the page, there are two buttons: "SIGN-UP FOR NEWSLETTERS" and "VIEW ALL NEWSLETTERS".

Questions?

